AWARD NUMBER: W81XWH-22-1-0751

TITLE: The Role of Genomic Diversity in Clinically High-Risk Prostate Cancer

PRINCIPAL INVESTIGATOR: Paul C Boutros

CONTRACTING ORGANIZATION: University of California, Los Angeles, CA

REPORT DATE: August 2023

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| **1. REPORT DATE** | **2. REPORT TYPE** | **3. DATES COVERED** |
|---|---|---|
| August 2023 | Annual | 01Aug2022-31Jul2023 |

**4. TITLE AND SUBTITLE**

The Role of Genomic Diversity in Clinically High-Risk Prostate Cancer

**5a. CONTRACT NUMBER**
W81XWH-22-1-0751

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Paul Boutros

E-Mail: pboutros@mednet.ucla.edu

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

The Regents of the University
of California, Los Angeles

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION / AVAILABILITY STATEMENT**

Approved for Public Release; Distribution Unlimited

**13. SUPPLEMENTARY NOTES**

None

**14. ABSTRACT**

**Hypothesis.** We hypothesize that the genomic diversity of localized prostate cancer predicts for metastatic spread of high-risk disease and is driven by specific oncogenes and tumor suppressors.

**Specific Aims.** We will first determine if genomic diversity predicts metastasis and lethality of high-risk prostate cancer (Specific Aim #1). We will then use a cohort of at least 750 prostate cancers with DNA whole-genome sequencing to identify specific driver genes that influence prostate cancer genomic diversity (Specific Aim #2).

**Study Design: Aim #1.** We have accrued 153 NCCN high-risk localized prostate cancers with high-coverage tumor and normal whole-genome sequencing. PGA and subclone number are the two measures of genomic diversity that most tightly predict metastasis of intermediate-risk prostate cancer. To rigorously test that association in high-risk disease we will perform high-coverage whole genome sequencing of tumor and blood normal samples from 100 NCCN high-risk localized prostate cancers with long-term outcome data. We will then determine whether these metrics individually and together predict metastasis and prostate cancer specific mortality.

**Study Design: Aim #2.** We have accumulated 750 localized prostate cancers with high-coverage tumor and normal whole-genome sequencing across all risk groups. Both PGA and subclone number vary widely within this cohort. To determine if specific driver somatic mutations influence these metrics of genomic diversity, we will use statistical modeling. We will quantify how each recurrent driver mutation influences the genomic diversity of a cancer, controlling for clinico-epidemiologic features like age and grade.

**15. SUBJECT TERMS**

None listed.

| **16. SECURITY CLASSIFICATION OF:** Unlimited | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON** |
|---|---|---|---|---|---|
| **a. REPORT** | **b. ABSTRACT** | **c. THIS PAGE** | | | USAMRDC |
| Unclassified | Unclassified | Unclassified | Unclassified | 13 | **19b. TELEPHONE NUMBER** *(include area code)* |

**TABLE OF CONTENTS**

## 1. INTRODUCTION:

**Prostate Cancer.** Prostate cancer is the most common non-skin cancer in men in the United States. It afflicts over 10% of men during the course of their lives. While many prostate cancers are non-lethal, others are extremely aggressive and kill men rapidly. While prostate cancers can stay localized to the prostate, many high-risk localized prostate cancers can spread or metastasize to other issues within the body. This metastatic spread of prostate cancers to other tissues can be hard to catch on initial diagnosis and therefore, affects clinical management and treatment. Interestingly, the genetics that lead to metastatic spread of high-risk prostate cancer are not well known. Therefore, understanding this process would potentially pave the way for improving cure of high-risk localized disease.

**Prostate Cancer Genomics.** Cancer at its root is a disease of the genome. Some of the three billion letters of DNA present in each cell will mutate, with one letter replacing another. These mutations can lead some cells to grow faster, or to be less capable of fixing mutations as they occur and leads to genomic diversity in each cancer. As a result, cancers slowly emerge over time. So, DNA mutations really underlie prostate cancer, and all cancers. However, this DNA information has not generally been helpful for prostate cancer patients. There are a few exceptions, like the wide-spread testing for *BRCA2* mutations. But while DNA sequencing has transformed many other types of cancer, it has had less impact on prostate cancer.

**Objective.** We hypothesize that tumors that evolve faster are more likely to metastasize and kill patients. We can have created algorithms to estimate the speed of evolution by looking at the patterns of mutations in a tumour. We will focus on patients with high-risk localized tumors – those at the highest risk of metastasis. First, we will use DNA sequencing and our algorithms to quantify whether evolutionary speed predicts for metastasis. Second, we will combine our new data with preexisting databases of prostate cancer DNA sequencing to try to understand why some tumors are evolving faster than others, and if specific cancer genes like *BRCA2* might explain this.

## 2. KEYWORDS:

Prostate
Cancer
Whole Genome Sequencing
Bioinformatics
DNA Mutations
Genomic Diversity
Germline
Big Data
Computational Algorithms
Databases
High-risk
Proportion of the Genome with a Copy Number Aberration
Cloud Computing

## 3. ACCOMPLISHMENTS:

a. **What were the major goals of the project?**

i. Specific Aim 1: Determine if genomic diversity predicts lethality of high-risk prostate cancer

1. Major Task 1: Obtain IRB and HRPO approval prior to initiating human data usage - COMPLETED

2. Major Task 2: Increase experimental cohort with whole genome sequencing – IN PROGRESS

ii. Specific Aim 2: Identify candidate drivers of genomic diversity

1. Major Task 1: Leverage existing datasets to identify biomarkers and drivers of genomic diversity – IN PROGRESS

b. **What was accomplished under these goals?**

**Specific Aim #1**

i. **Overview. Study Design: Aim #1.** We have accrued 153 NCCN high-risk localized prostate cancers with high-coverage tumor and normal whole-genome sequencing. PGA and subclone number are the two measures of genomic diversity that most tightly predict metastasis of intermediate-risk prostate cancer. To rigorously test that association in high-risk disease we will perform high-coverage whole genome sequencing of tumor and blood normal samples from 100 NCCN high-risk localized prostate cancers with long-term outcome data. We will then determine whether these metrics individually and together predict metastasis and prostate cancer specific mortality.

ii. **Major Activity #1:** IRB and HRPO approval have been completed.

iii. **Major Activity #2:** We have initiated sequencing and analysis of three cohorts. First, in collaboration with Dr. Robert Reiter at UCLA, we have identified a cohort of high-risk prostate cancer patients with fresh frozen Radical prostatectomy and blood samples. DNA was extracted from a subset of these patients and subject to deep whole genome sequencing to verify quality and usability of the cohort before proceeding with comprehensive sequencing of the full sample set. Tumor-normal sample pairs are critical to understanding differences in tumor biology from normal cells. Sequencing of both blood and tumor samples was analyzed using a standardized quality control workflow to assess read quality, sequence length distribution, and a variety of other metrics (**Figure 1**). Sequencing coverage was assessed by sample and chromosome (**Figure 2**). Pathologic annotation of tumor samples is currently in progress. Sequencing data is being processed through standardized bioinformatic pipelines to quantify the full range of information from a DNA study, including somatic single nucleotide variation, copy number and driver mutations.

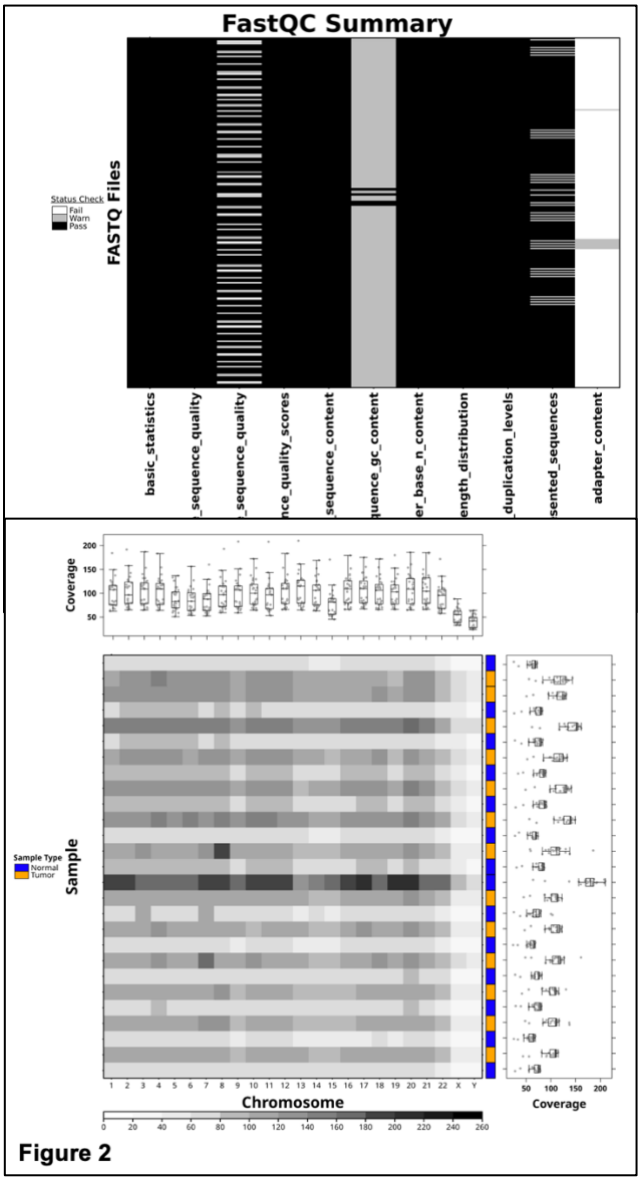

Figure 2

Second, as a, complementary cohort, we collaborated with Dr. Stanley Liu, a radiation oncologist. Dr. Liu's team has identified several high-risk patients with atypically aggressive clinical behavior. We have sequenced a set of Gleason score 9 and 10 patient tumors. These tumors represent a variety of metastatic potentials, with varying somatic mutational profiles. We have recently completed deep whole genome sequencing of these 10 samples alongside paired normal blood samples. Using a standardized quality control workflow, we assessed read quality on a variety of metrics such as per-base sequence quality, sequence length distribution, and sequence duplication. Additionally, we verified depth of coverage across samples and chromosomes. (**Figure 3**). All sequencing runs meet our quality control standards. We are currently processing these samples using standardized bioinformatic pipelines focused on calling somatic, germline and structural variants. This will develop a unique resource to study prostate cancers in high-risk patients with varying metastatic potential and identify determinants of aggression.

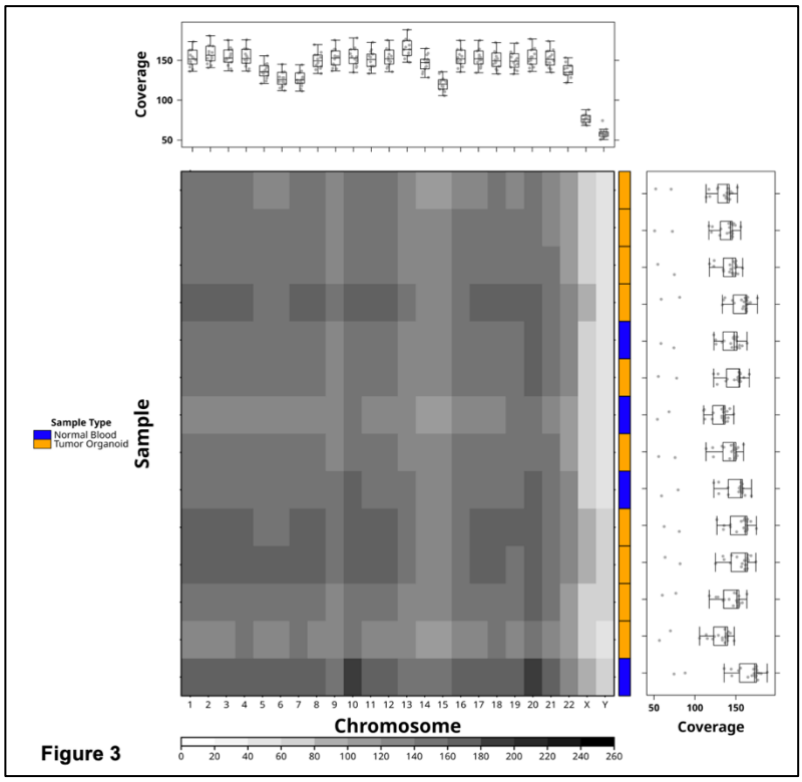

Figure 3

Third and finally, cancer cell lines are vital tools in research as they allow us to study the behavior and genetics of tumors in a controlled environment. Unlike in other cancer types, there are relatively few cell lines that are commonly used in prostate cancer research, most of which have not been characterized using DNA whole genome sequencing. Further, almost all prostate cancer cell-lines derive from high-risk tumours. By sequencing the most used prostate cancer cell lines, we aim to identify the shared genomic features of these model systems and identify missing genomic diversity that is present in primary tumors as part of our high-risk cohort. We have completed deep whole genome sequencing on the most used prostate cancer cell lines and processed them using a systematic bioinformatics pipeline. We are examining model-specific features that may define aggression using the full range of information from a DNA study, including total mutational load, mutational signatures, driver mutations, and evolutionary features like mutational timing. Using this unique resource, we are currently exploring copy number

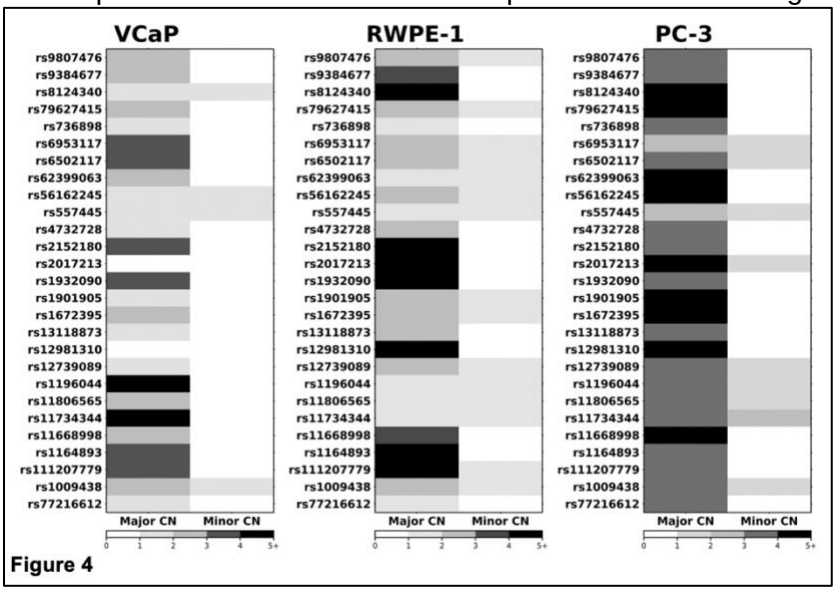

Figure 4

amplification of 30 germline single nucleotide polymorphisms (**Figure 4**) between cell lines.

**Specific Aim #2**

i. **Overview. Study Design: Aim #2.** We have accumulated 750 localized prostate cancers with high-coverage tumor and normal whole-genome sequencing across all risk groups. Both PGA and subclone number vary widely within this cohort. To determine if specific driver somatic mutations influence these metrics of genomic diversity, we will use statistical modeling. We will quantify how each recurrent driver mutation influences the genomic diversity of a cancer, controlling for clinico-epidemiologic features like age and grade.

ii. **Major Activity #1:** We assembled a set of 768 whole-genome sequenced tumour-normal pairs from prostate cancers representing the full spectrum of grade, with extensive clinical characterization. All samples were processed using a systematic bioinformatics pipeline, making the resulting molecular data a large-scale resource for our community, along with its detailed and systematized clinical information. This collection is, to our knowledge, the largest whole genome analysis of a single cancer type to date. Using this unique resource, we created the first exhaustive compendium of driver aberrations in prostate cancer. Our compendium contains 223 recurrently mutated genomics regions: comparable to the 299 identified in all 30 cancer types in TCGA pan-cancer studies. The reason we found so many new drivers: over 37% of prostate cancer drivers are silent to targeted sequencing assays because they involve structural variation. A salient example is *FOXA1*, which is mutated in protein-coding regions in 5.8% of cases, but in other ways in 10.1% more cases, with transcriptional consequences. Moreover, we evaluated the evolutionary history of these drivers using subclonal reconstruction techniques, statistical analyses and survival analyses. We found that while most prostate cancer mutations occur subclonally, most driver mutations occur clonally, and happen very early in tumor development. We identify 16 specific driver regions that appear to define lethal, high-grade tumors, including *BRCA2* and *MYC*. We then identify specific drivers associated with each of the clinico-epidemiologic factor used in prostate cancer management. These largely do not overlap, matching the clinical observation that these are independent prognostic features. Further, all prognosis-associated mutations preferentially occur clonally. Finally, we show, to our knowledge for the first time in any cancer type, that the evolutionary timing of a mutation matters: the same mutation occurring clonally and subclonally have different clinical consequences. Taken together these multiple lines of evolutionary inference data strongly support a new model for the evolution of prostate cancer grade: both low- and high- grade tumors evolve from a common mutagenic field, with the acquisition of specific driver mutations like *BRCA2* leading to clonal expansion of aggressive subclones. **Figure 1** below.

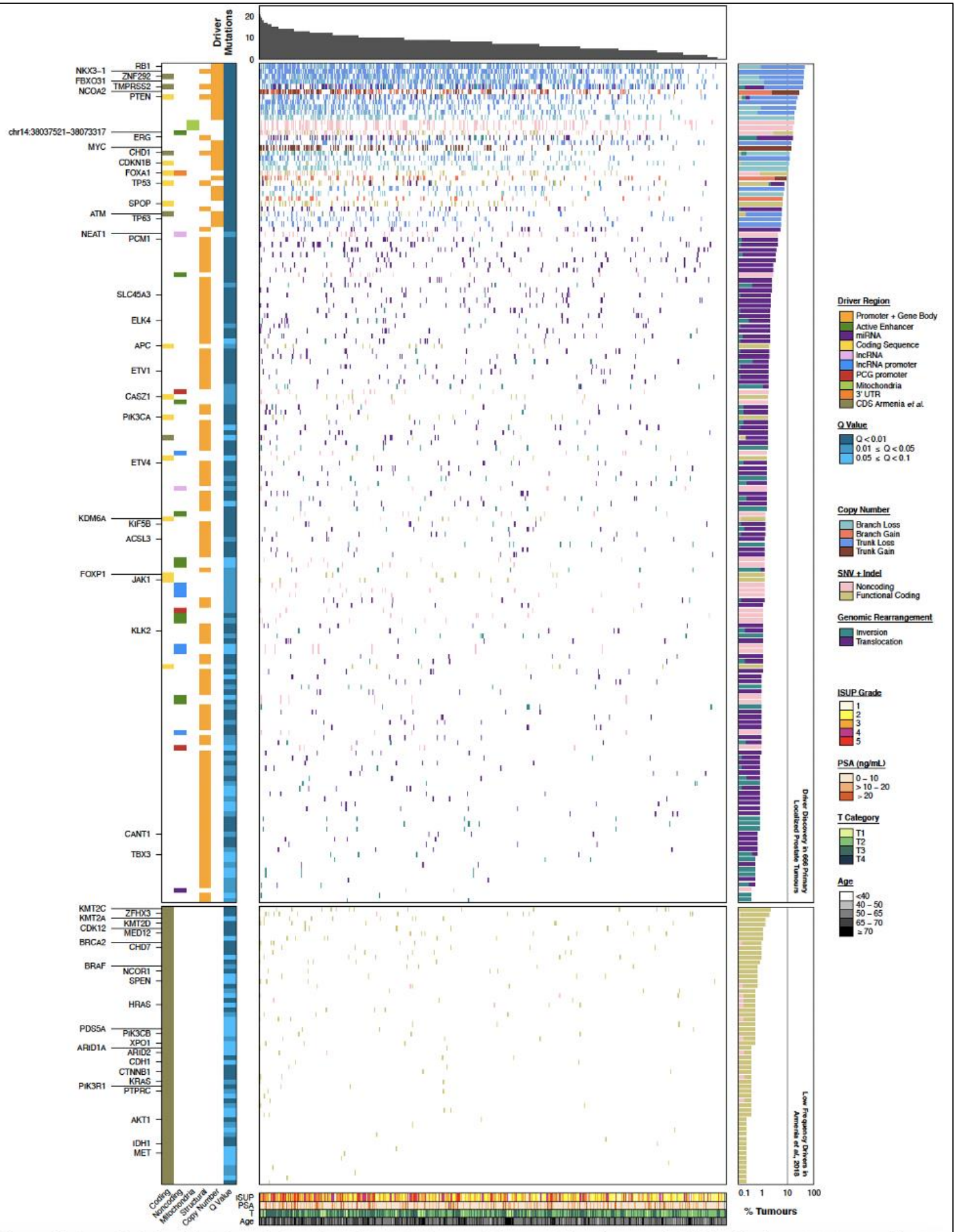


**Figure 1 | Somatic Driver Mutations in Localized Prostate Cancer.** Driver mutation discovery in 666 localized prostate tumors. The top barplot shows the distribution of the number of drivers in patients; the covariates on the left show the region type and statistical significance from ActiveDriverWGS and GISTIC. The top heatmap shows drivers found in this study (rows) for each patient (columns). The bottom heatmap shows drivers found in an exome meta-analysis. Heatmaps are colored by mutation type. Right barplot shows the number of drivers per patient. Bottom covariate bars show clinical features of patients. Gene labels on the left are for rows identified by tick-marks on the axis.

c. **What opportunities for training and professional development has the project provided?**

i. While the project was not specifically intended to provide training, our bioinformatician (Yash Patel) working on this project have been involved in mentorship of PhD students and fellow staff members throughout the course of their work on this project. Additionally, all our staff consistently present their work during our weekly lab meetings, local seminars (including from our Prostate SPORE and Cancer Center) and several other venues.

d. **How were the results disseminated to communities of interest?**

i. Nothing to Report

e. **What do you plan to do during the next reporting period to accomplish the goals?**

i. **Specific Aim 1:** Determine if genomic diversity predicts lethality of high-risk prostate cancer

1. **Major Activity 1:** Obtain IRB and HRPO approval prior to initiating human data usage - COMPLETED

2. **Major Activity 2:** Increase experimental cohort with whole genome sequencing – In year 2, we plan to increase our sequencing cohort of high-risk localized prostate cancer samples (as outlined in our Statement of Work) with 45 new samples. As new sequencing is completed, we will perform bioinformatics analysis to quantify the association of metastasis with genomic features. We will perform subclonal reconstruction and assess metrics of genomic diversity including genomic instability and tumor clonality.

ii. **Specific Aim 2:** Identify candidate drivers of genomic diversity

1. **Major Activity 1:** Leverage existing datasets to identify biomarkers and drivers of genomic diversity – In year 2, we will compare our sequencing data to existing datasets to find biomarkers and drivers of genomic diversity (outlined as ongoing through the whole project period in our Statement of Work). Furthermore, as a conclusion of this project, we plan to integrate our sequencing data and dataset analysis with clinical features including ISUP Grade and T-category (size and extent), as well as construct models that predict an individual tumor's diversity.

**4. IMPACT:**

a. **What was the impact on the development of the principal discipline(s) of the project?**

i. **Short-term Impact.** Our work will have three major short-term impacts. First, we will determine if genomic diversity is a robust biomarker for metastatic spread of high-risk localized disease. If tumors with high genomic diversity show a substantially elevated risk of progression, then treatment intensification with adjuvant hormonal therapy might be considered. Similarly, if high-risk tumors with low genomic diversity were at low risk for metastasis, then patients receiving radiotherapy (who are typically more elderly) might be considered for treatment deintensification by eliminating adjuvant hormonal therapy. Second, this proposal will also generate a key resource that is currently missing: a large dataset of clinically well-annotated high-risk localized patients with deep whole-genome sequencing of tumor and normal, and associated long-term clinical information. Indeed, the majority of molecular datasets in this space are limited to the transcriptome or small targeted germline or somatic sequencing panels. There will thus be immediate resource-value to and impact from this study, even if our hypothesis is not supported by it. Third, we and others have identified several hundred prostate cancer driver genes. These genes,

when somatically mutated, drive the initiation and progression of prostate cancer. For most, we have only limited understanding of either their mechanistic functions or the ways in which they shape the subsequent evolutionary trajectory of tumors. This study will produce an atlas outlining exactly which driver genes influence which aspects of prostate cancer genomic evolution. This will inform mechanistic and model development strategies broadly.

ii. **PCRP Overarching Challenges.** This study focuses on outcomes of men with high- and very high-risk localized prostate cancer and in defining the biology of this disease state. It studies determinants of metastasis and lethality, identifying the specific individual driver genes that promote the evolution of lethal disease. This work therefore addresses directly two of the PCRP over-arching challenges.

b. **What was the impact on other disciplines?**

i. **Long-term Impact: Biology.** In the longer-term, understanding how and why lethal prostate cancers evolve the capacity to escape the gland and evolve treatment resistance is fundamental to the design of advanced management strategies. This work has two key long-term impacts. First, it will provide fundamental biological understanding. We do not yet understand why certain prostate tumors escape the gland and become established (primarily) in lymph nodes and bone. There are likely a range of factors underlying this behavior, and it appears genomic diversity is one of the best predictors of it. By determining if the relationship between diversity and metastasis holds for all localized tumors or is stronger for lower-grade or higher-grade ones we will learn a fundamental aspect of prostate tumor development. Similarly, by identifying which specific gene mutations (if any!) drive increasing genomic diversity we will create avenues for designing evolutionarily-aware interventions not currently possible. For example, if we can predict which high-risk prostate cancers are likely to be generating occult metastatic spread with subclonal defects in homologous recombination, we could consider DNA damage-directed therapies like inhibition of PARP in patients who do not have germline or clonal *BRCA2* mutations. By contrast, if no such mutations exist, this work would redirect efforts towards features beyond cancer driver genes, like the tumor proteome or its microenvironment.

ii. **Long-Term Impact: Biomarkers.** A majority of men with high-grade localized prostate cancer will experience metastatic spread of their disease. Strong trial data has led to patients managed surgically to typically not receive adjuvant hormone therapy, while those treated radiotherapeutically do. Our work will establish if genomic diversity is a strong predictor, or even a requirement for metastatic spread of high-risk disease. If it is, then the long-term impact would be substantial: genomic biomarkers might provide adjuncts to existing clinical, radiologic, pathologic and transcriptomic features to support risk-stratification and design clinical trials to improve cure rates and reduce over-treatment of high-risk localized prostate cancer.

c. **What was the impact on technology transfer?**

i. Nothing to report

d. **What was the impact on society beyond science and technology?**

i. Nothing to report

## 5. CHANGES/PROBLEMS:

a. **Changes in approach and reasons for change**

i. Nothing to report

b. **Actual or anticipated problems or delays and actions or plans to resolve them**

   i. Nothing to report

c. **Changes that had a significant impact on expenditures**

   i. We have started sequencing in year 1, however, we intent to ramp up sequencing efforts in year 2.

d. **Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents**

   i. Nothing to report

e. **Significant changes in use or care of human subjects**

   i. Nothing to report

f. **Significant changes in use or care of vertebrate animals.**

   i. Not applicable

g. **Significant changes in use of biohazards and/or select agents**

   i. Not applicable

## 6. PRODUCTS:

- **Publications, conference papers, and presentations**
  - **"The Germline and Somatic Origins of Prostate Cancer Heterogeneity"** Takafumi N. Yamaguchi, Kathleen E. Houlahan, Helen Zhu, Natalie Kurganovs, Julie Livingstone, Natalie S. Fox, Jiapei Yuan, Jocelyn Sietsma Penington, Chol-Hee Jung, Tommer Schwarz, Weerachai Jaratlerdsiri, Job van Riet, Peter Georgeson, Stefano Mangiola, Kodi Taraszka, Robert Lesurf, Jue Jiang, Ken Chow, Lawrence E. Heisler, Yu-Jia Shiah, Susmita G Ramanand, Michael J. Clarkson, Anne Nguyen, Shadrielle Melijah G. Espiritu, Ryan Stuchbery, Richard Jovelin, Vincent Huang, Connor Bell, Edward O'Connor, Patrick J. McCoy, Christopher M. Lalansingh, Marek Cmero, Adriana Salcedo, Eva K.F. Chan, Lydia Y. Liu, Phillip D. Stricker, Vinayak Bhandari, Riana M.S. Bornman, Dorota H.S. Sendorek, Andrew Lonie, Stephenie D. Prokopec, Michael Fraser, Justin S. Peters, Adrien Foucal, Shingai B.A. Mutambirwa, Lachlan Mcintosh, Michèle Orain, Matthew Wakefield, Valérie Picard, Daniel J. Park, Hélène Hovington, Michael Kerger, Alain Bergeron, Veronica Sabelnykov, Ji-Heui Seo, Mark M. Pomerantz, Noah Zaitlen, Sebastian M. Waszak, Alexander Gusev, Louis Lacombe, Yves Frade, Andrew Ryan, Amar U. Kishan, Martijn P. Lolkema, Joachim Weischenfeldt, Bernard Tetu, Anthony J. Costello, Vanessa M. Hayes, Rayjean J. Hung, Housheng H. He, John D. McPherson, Bogdan Pasaniuc, Theodorus van der Kwast, Anthony T. Papenfuss, Matthew L. Freedman, Bernard J. Pope, Robert G. Bristow, Ram S. Mani, Niall M. Corcoran, Juri Reimand, Christopher M. Hovens, Paul C. Boutros – UNDER REVIEW, Cancer Discovery
  - **"Colibactin Exerts Androgen-Dependent and -Independent Effects on Prostate Cancer"** Raag Agrawal, Sarah Al-Hiyari, Rupert Hugh-White, Robert Hromas, Yash Patel, Elizabeth A Williamson, Mohammed F.E. Mootor, Alfredo E Gonzales, Jianmin Fu, Roni Haas, Madison Jordan, Brian L Wickes, Ghouse Mohammed, Mao Tian, Christian Jobin, Takafumi N Yamaguchi, Seth B Herzon, Daniel R Semlow, Paul C Boutros, Michael A. Liss – IN PREPARATION
- **Website(s) or other Internet site(s)**

- Nothing to report

- **Technologies or techniques**
  - Nothing to report
- **Inventions, patent applications, and/or licenses**
  - Nothing to report
- **Other Products**
  - Nothing to report

## 7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

- **What individuals have worked on the project?**

| | |
|---|---|
| Name: | Paul Boutros |
| Project Role: | PI |
| Researcher Identifier (e.g. ORCID ID): | 0000-0003-0553-7520 |
| Nearest person month worked: | 1 |
| Contribution to Project: | Dr. Boutros serves as the PI of this project. |
| Funding Support: | Dr. Boutros receives support from this project. |

| | |
|---|---|
| Name: | Yash Patel |
| Project Role: | Programmer Analyst III (Bioinformatician) |
| Researcher Identifier (e.g. ORCID ID): | 0000-0003-3113-7010 |
| Nearest person month worked: | 2.5 |
| Contribution to Project: | Mr. Patel serves as a bioinformatician for this project |
| Funding Support: | Yash receives additional funding from UCLA internal grants and departmental resources. |

- **Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?**
  - Since receipt of this award, Dr. Boutros has received several other awards including: DoD W81XWH2210631, DoD W81XWH2210569, NIH U54HG012517, NIH U2CCA271894, NIH R01CA272678, and NIH R01CA270108. Due to this, we have reduced some of Dr. Boutros' effort on grants that allowed reduction. None of these have overlap with the present grant.
- **What other organizations were involved as partners?**

- Nothing to Report

## 8. SPECIAL REPORTING REQUIREMENTS

- Nothing to report

## 9. APPENDICES

- Nothing to report